Professor Christodoulos A. Floudas

Stephen C. Macaleer '63 Professor in Engineering and Applied Science,
Department of Chemical Engineering, Engineering Quadrangle,
Princeton University, Princeton NJ 08544
Email: floudas@titan.princeton.edu
http://titan.princeton.edu

Secondary Structure Prediction Problems

1. Purely α – helical protein

Problem: Predict the secondary structure of the following purely α – helical protein (PDB: 1R69, Number of residues = 63)

SISSRVKSKRIQLGLNQAELAQKVGTTQQSIEQLENGKTKRPRFLPELASALGVSV
DWLLNGT

2. Purely β – protein

Problem: Predict the secondary structure of the following purely β – protein (PDB: 1TEN, Number of residues: 89)

LDAPSQIEVKDVTDTTALITWFKPLAEIDGIELTYGIKDVPGDRTTIDLTEDENQYS
IGNLKPDTEYEVSLISRRGDMSSNPAKETFTT

3. Mixed α/ β protein

Problem: Predict the secondary structure of the following mixed α/ β protein (PDB: 121P, Number of residues = 166)

MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDI
LDTAGQEEYSAMRDQYMRTGEGFLCVFAINNTKSFEDIHQYREQIKRVKDSDDV
PMVLVGNKCDLAARTVESRQAQDLARSYGIPYIETSAKTRQGVEDAFYTLVREI
RQH

Hint: EVASEC (http://cubic.bioc.columbia.edu/eva/doc/intro_sec.html) is a portal, which automatically analyses protein secondary structure prediction servers in 'real time'. A number of online secondary structure prediction methods can be found here. Some of the top – most servers from here can be used for secondary structure prediction. The predictions from the servers can be used in a consensus evaluation of the final secondary structure of the protein.

β – sheet Topology Prediction Problems

1. BPTI is a small globular protein found in many tissues of the human body. BPTI inhibits several of the serine protease proteins such as trypsin, kallikrein, chymotrypsin and plasmin, and is a part of the family of serine protease inhibitors. These proteins usually have conserved cysteine residues that participate in the formation of disulphide bonds.

   Predict the sheet topology for the BPTI protein presented below. The actual secondary structure has also been presented:

   RPDFCLEPPYTGPCKARIIRYFYNAKAGLCQTFVYGGCRAKRNNFKSAED CMRTCGGA

   The secondary structure is:

   Helix 1:   3 to 6
   Strand 1:   18 to 24
   Strand 2:   29 to 35
   Strand 3:   42 to 43
   Helix 2:   48 to 55

   Predict the β – sheet topology for BPTI, using the strand – wise and the residue – wise hydrophobicity maximization Integer programming model (ILP).

Tertiary Structure Prediction Problems

1. The prediction of the tertiary structure of a protein, given its amino acid sequence (primary sequence) is considered as one of the "holy grails" of computational biology. Numerous methods have been suggested towards this end, including homology based, fold recognition based and *ab initio* prediction methods.

   The TPR2A domain (PDB: 1elr, Number of residues = 128) of the HOP complex is an important chaperone protein, which is a critical element in the assembly of the Hsp70-Hsp90 multi-chaperone machine in the body. The amino acid of this purely α – helical domain is:

   GKQALKEKELGNDAYKKKDFDTALKHYDKAKELDPTNMTYITNQAAVY FEKGDYNKCRELCEKAIEVGRENREDYRQIAKAYARIGNSYFKEEKYKD AIHFYNKSLAEHRTPDVLKKCQQAEKILKEQ

   Servers like SAM (http://www.soe.ucsc.edu/research/compbio/SAM_T06/T06-query.html) and 3DPro (http://scratch.proteomics.ics.uci.edu/) provide online

facility to submit amino acid sequences to get back the final 3-d structures of proteins. The response is provided in PDB format, and can be viewed using commonly available molecular viewers.

Predict the structure of 1ELR, and compare the predicted structures with the actual native structure of the protein. The native structure of 1ELR can be got from the Protein Data Bank (http://www.rcsb.org/pdb/home/home.do). One of the common ways to compare structures is to evaluate the root mean squared deviation (RMSD) between the two structures.

Hint: In order to compare the outputs, the outputs can be viewed on the free, open – source molecular viewer PYMOL (available at www.pymol.org). The older, freely available version of PYMOL can execute RMSD evaluations. After installing PYMOL, run the pymol executable. Load the predicted protein structure, along with the native protein structure onto pymol. Once this is done, you can click on the action button corresponding to either one of them (the 'A' button next to the structure name), and align it to the other structure. The RMSD value between the two shall show up on the supporting text screen.